AD696232

FTD-MT-24-133-69

FOREIGN TECHNOLOGY DIVISION

ALL-UNION CONFERENCE ON INFORMATION RETRIEVAL SYSTEMS AND AUTOMATIC PROCESSING OF SCIENTIFIC AND TECHNICAL INFORMATION, 3rd, MOSCOW, 1966, TRANSACTIONS

(Selected Articles)

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

66

This document is a machine translation of Russian text which has been processed by the AN/GSQ-16(XW-2) Machine Translator, owned and operated by the United States Air Force. The machine output has been post-edited to correct for major ambiguities of meaning, words missing from the machine's dictionary, and words out of the context of meaning. The sentence word order has been partially rearranged for readability. The content of this translation does not indicate editorial accuracy, nor does it indicate USAF approval or disapproval of the material translated.

EDITED MACHINE TRANSLATION

ALL-UNION CONFERENCE ON INFORMATION RETRIEVAL SYSTEMS AND AUTOMATIC PROCESSING OF SCIENTIFIC AND TECHNICAL INFORMATION, 3rd, MOSCOW, 1966, TRANSACTIONS

English pages: 53

Source: Vsesoyuznoy Konferentsii po Informatsionno-Poiskovym Sistemam i Avtomatizirovannoy Obrabotke Nauchno-Tekhnicheskoy Informatsii 3D, Moscow, 1967, Trudy, Vol. 4, pp. 3-4, 25-38, 203-206, 222-229, 300-305.

THIS TRANSLATION IS A RENDITION OF THE ORIGINAL FOREIGN TEXT WITHOUT ANY ANALYTICAL OR EDITORIAL COMMENT. STATEMENTS OR THEORIES ADVOCATED OR IMPLIED ARE THOSE OF THE SOURCE AND DO NOT NECESSARILY REFLECT THE POSITION OR OPINION OF THE FOREIGN TECHNOLOGY DIVISION.	PREPARED BY: TRANSLATION DIVISION FOREIGN TECHNOLOGY DIVISION WP-AFB, OHIO.

TABLE OF CONTENTS

U. S. BOARD ON GEOGRAPHIC NAMES TRANSLITERATION SYSTEM

Block	Italic	Transliteration	Block	Italic	Transliteration
А а	*А а*	A, a	Р р	*Р р*	R, r
Б б	*Б б*	B, b	С с	*С с*	S, s
В в	*В в*	V, v	Т т	*Т т*	T, t
Г г	*Г г*	G, g	У у	*У у*	U, u
Д д	*Д д*	D, d	Ф ф	*Ф ф*	F, f
Е е	*Е е*	Ye, ye; E, e*	Х х	*Х х*	Kh, kh
Ж ж	*Ж ж*	Zh, zh	Ц ц	*Ц ц*	Ts, ts
З з	*З з*	Z, z	Ч ч	*Ч ч*	Ch, ch
И и	*И и*	I, i	Ш ш	*Ш ш*	Sh, sh
Й й	*Й й*	Y, y	Щ щ	*Щ щ*	Shch, shch
К к	*К к*	K, k	Ъ ъ	*Ъ ъ*	"
Л л	*Л л*	L, l	Ы ы	*Ы ы*	Y, y
М м	*М м*	M, m	Ь ь	*Ь ь*	'
Н н	*Н н*	N, n	Э э	*Э э*	E, e
О о	*О о*	O, o	Ю ю	*Ю ю*	Yu, yu
П п	*П п*	P, p	Я я	*Я я*	Ya, ya

* ye initially, after vowels, and after ъ, ь; e elsewhere. When written as ё in Russian, transliterate as yё or ё. The use of diacritical marks is preferred, but such marks may be omitted when expediency dictates.

PREFACE

The Party and Government are giving their constant attention to the development and improvement of systems of the information service. The special resolution adopted by the Council of Ministers of the USSR in November 1966 is directed towards implementing solutions to the problems enacted by the 23rd Congress of the CPSU on creation of government-wide highly effective scientific-technical information service.

Being based on contemporary technical means for acquisition, processing, investigation and delivery of information data, and automation of information processes, the State system of scientific information is directed to provide timely information about achievements in domestic and foreign science and technology.

At present in many organizations of the country scientific research work is being conducted in the area of automation of information processes. The level and state of this work require defined organization and coordination. This is especially important since scientific and technical information is such a many-faceted branch of science and technology, embracing all branches of the national economy, that its further effective development would be inconceivable without coordination of the work and exchange of the experience of all specialists occupied with automation of information processes.

The Third All-Union Conference on information retrieval systems and automated processing of scientific and technical information, conducted from 19 through 22 December 1966 in Moscow, provided results of scientific and technical information activity of scientists and specialists and outlined the way for the fullest and fastest fulfillment of problems set by the Party and Government in the area of scientific and technical information.

The work of conference involved the participation of 1150 specialists representing different ministries, departments, information organs, scientific research and design organizations, industrial enterprises and establishments of the country. At the conference 220 reports were presented.

The present Transactions contain material presented at the plenary sessions of conference, as well as at meetings of the separate sections.

For the convenience of the readers, the Transactions are being issued in four volumes.

The first volume "Information Retrieval Systems" presents the results of research and development in the areas: semantic systems of investigation of scientific and technical literature in large data bases, with automatic translation from a natural into a formalized language; automated systems of factographic facilities based on the use of information languages of the natural sciences; information retrieval systems for industrial enterprises and establishments.

The same volume is devoted to other questions connected with the creation and introduction of automated information retrieval systems of different classes and assignment for processing large as well as small volumes of scientific and technical information.

The second volume, Semiotic Problems of Automatic Data Processing" presents material dedicated to: development of problems

of the connection between syntactic and semantic properties of language systems; investigation of the natural and formalized languages of science and technology in connection with problems of storage and retrieval of information; questions of automatic processing of tests for creating operational systems of machine indexing, abstracting and translation of texts; research in the area of creation of special programming languages and translators from them for machine processing of texts.

The third volume, "Automatic Reading Devices" presents results of research and development of different methods and devices for automatic identification of typographical and typewritten symbols. Special attention is alloted here to automatic reading machines, allowing automated computer input of masses of scientific and technical, statistical and economic information to ETsVM [electronic digital computers].

The fourth volume, "Technical Devices for the Information Service and Immediate Reproduction Techniques" presents works related to research and development of: technical devices for preparation and input of alphanumeric information into a computer and high speed output devices of textual information, and also output devices with many characters and high quality print; specialized memory units for information systems, possessing internal logic and the possibility of storage of large volumes of information, including photoscopic, associative, with internal logic; retrieval devices on continuous carriers (microfilms) and discrete carriers (microphoto cards, magnetic cards, etc.); technical means for mechanization and automation of all stages of information processes.

A considerable place in this volume is given to questions on organization of industry and improvement of technological processes on output of urgent information publications with wide application of methods of the classical printing industry and reproduction technology, the use of latest models of typesetting typewriters with

stored control for manufacture of original mock-ups suitable for direct reproduction, application of xerographic, electronic equipment for urgent manufacture of printing forms, and also questions of the use of high-speed cylinder and offset machines for printing information.

Material is given on the use of highly productive brochure equipment.

QUESTIONS OF THE AUTOMATION OF INPUT, OUTPUT, AND DATA PROCESSING ON COMPUTERS IN SYSTEMS OF INFORMATION SERVICE

O. V. Mamontov, M. L. Avrukh, and
V. A. Kal'manson

The examination and processing of enormous masses of information, their systematization and organization of storage, interrogation retrieval and output of the latter into definite form are characteristic for the development of contemporary science, technology and other branches of activity of man. A similar problem is encountered by scientists studying space, designers creating new models of aircraft and rockets, economists and medics, chemists and biologists; this problem acquired specific importance in the resolution of problems of scientific and technical information. Enormous masses of processed information made the problem of the automation of information processes very urgent with the help of methods of computer technology.

It is obvious that in the near future not one complicated problem of information service will be solved without the application of computers and means of information technology.

Various problems solvable in systems of information service can be conditionally referred to several different directions. The most important of them are problems connected with the creation

of automated information retrieval systems (IRS) of different types and the problem on machine preparation of information publications; a special group is comprised or problems connected with new developments and scientific investigations.

All these problems were solved earlier in accordance with produced rules, methods and technical possibilities. Complicated problems were simplified, and solutions were found approximately or empirically. The use of computers considerably expands the logical and calculating possibilities. With their help such sides of various information processes which were inaccessible for accurate analysis by other means are investigated.

Results of works conducted in All-Union Institute of Scientific and Technical Information (VINITI) on the creation of automated IRS are well-known. With these results it was possible to become familiar with international exhibitions of "Inforga-65" and "Interorgtekhnika-66"; they are widely discussed in literature.

Works on the use of computers in the preparation of different information publications were started in VINITI just in 1966.

Engineers A. B. Potapovskiy and M. A. Zaytsev created programs, developed and accomplished the technology of the output of permuted indexes. The whole process from the input of initial information on punched cards to the obtaining of an offset form from a tabulogram [Translator's Note: this term is not verified] print-out of the output computer device will be fully automated. Results of the conducted experiment will allow in the very near future turning to the automated output of signal information with the help of computers.

Successfully being conducted by Ya. I. Posternak and V. P. Korobkov is the development of algorithms, programs, and technology of the application of computers for the automated preparation of dictionaries, directories, information, and signal literature and other information publications.

We assume that the development of effective algorithms, improved technology and the use of information-logic system with immediate memory of the order of 30-60 thousand words considerably approaches the solution to this problem. The experiment performed in VINITI permitted estimating the effectiveness of computer application, generalizing technical requirements for information-logic systems and definitizing directions of investigations.

Peculiarities of Applied Computers

To solve information problems at present, as a rule, there are being used universal serial computers of the "Ural" and "Minsk" type and other computers accessible for consumers of means of computer technology. All these machines have definite technical characteristics and inherent peculiarities, which determine the possibility of their application for the automation of information processes:

1) comparatively well-developed instruction system containing several dozens of operations;

2) small capacity of the immediate memory unit (IMU) which consists of 4096 cells;

3) possibility of the fulfillment of operations only on words stored in the IMU;

4) speed of fulfillment of operations in the arithmetical memory unit (MU) of 5-6 thousand operations per second;

5) capacity of the external MU, calculated in the tens and hundreds of thousand words;

6) low speed of exchange of numbers between operational and external MU – 1000-3000 words per second, the exchange is produced with stop in the calculations;

7) relatively low reliability of operation of external MU and input and output devices;

8) data input into the machine only from punched cards and punched tapes at a rate of a hundred words per second with a stop of the machine;

9) data output from the machine on the perforator at a rate of several words per second and print-out on paper tape at a rate of 3-8 lines per second;

10) impossibility of the parallel operation of devices and the absence of a system of multiprogram work and priority;

11) insufficiently high readability of texts printed by output device;

12) impossibility of print-out of chemical and mathematical formulas, special signs and various alphabets.

By their technical characteristics the machines appear to be the most adapted for solving such problems for which the decision algorithm realized in the program possesses small connectivity. The program of calculations contains a comparatively small number of instructions, which together with the initial data are completely placed into the IMU, and in the process of calculations it is not necessary to address the external storage units. When the quantity of operations in the calculations exceeds the volume of the immediate memory unit, then it is desirable that this occurs not due to the input of new parts of the program stored in the external storage units but owing to cyclical repetition of sections of the program found in the IMU. The access to external storage units at which stop of the computer occurs lowers the total speed of the calculating process and increases the probability of the appearance of errors because errors appearing with the rewriting of masses of information, and therefore it is important to reduce the excess of initial information introduced into the machine and use algorithms, which allow creating the most compact cyclical programs.

To evaluate the effectiveness of application of existing computers, general analysis of information problems and applied decision algorithms is important.

Peculiarities of Processed Information

In order to solve problem of information service, it is necessary to process large masses of different documents, the

information in which is contained mainly in literal form; furthermore, in documents there can be encountered figures, mathematical and chemical symbols and formulas, and also different diagrams and figures.

Every report entering into the document consists of separate words of variable length. The word minimum in length contains one symbol and the maximum – more than two dozen symbols.

Analysis shows that in the Russian language words consisting of 4-12 letters are the most widespread. For each word of variable length it is usually necessary to select one cell of the computer memory. Words of great length occupy several cells. Because of the fact that the length of the processed words is variable, in a number of cases it is necessary to use criteria of the beginning and end of the report and also criteria of the separation of separate words.

The peculiarity of operable numbers is their integral form.

General Characteristic of Processing Algorithms

Algorithms of data processing, which are used in the solution of problems of information service, have a number of distinctive peculiarities:

1) the presence of a large number of logical operations; in certain problems their quantity reaches 80-90% of the total number of instructions in the program;

2) great proportion of sending operations, which occupy second place in use in the programs;

3) cyclic recurrence of processing programs, i.e., recurrence of separate divisions of the program, groups of instructions and separate instructions, which are fulfilled on different words.

Thus, consideration of peculiarities of information and decision algorithms shows that applicable types of Soviet digital

computers with respect to their parameters, characteristics and reliability are insufficiently adapted to the solution of information problems.

The necessity to apply information manually on punched tapes and punched cards, the low speed of input and output of data, small volume of the IMU and external storage units, low speeds of rewrite from external MU and low reliability of their operation considerably lower the effectiveness of the application of universal computers in the solution of information problems connected with the input and output of great masses of reports. The processing of enormous masses with respect to comparatively simple programs with a relatively small variety of operations does not permit using wide possibilities embodied in the instruction systems of contemporary computers for logical data processing. In the process of the operation it is necessary constantly to have access to external storage units. In connection with the fact that rewriting proceeds at a low rate and halt of the machine is obligatory, the effectiveness of the application of high speed immediate and arithmetical MU decreases.

Memory units are used irrationally, since it is necessary to store great information masses consisting of words of variable length.

Thus, because of the insufficient adaptability of existing universal computers to solving information problems there is great urgency in the work on the creation of information-logic systems with various input devices, specialized MU and high speed output printers, which allow issuing information in the form of figures and letters of different outline, formulas and graphs, and externally this information should differ little from the typographical set.

Thus, questions connected with input, storage and output in the solving of problems of information service on computers acquire specific importance.

Problem Information Input

In connection with the growth in volume of storable information, the increase in high-speed operation and complication of functions of contemporary computers, the process information input becomes essentially a bottleneck, inasmuch as existing input devices from punched tapes, punched cards and other forms of carriers cannot provide the necessary speed and reliability of input. It is practically impossible to construct any developed information system with an extensive storehouse of information on the basis of the manual transfer of documents from the sheet to punched carriers. The process of manual transfer of the initial document on punched cards or punched tapes should be completely excluded, otherwise information data will become obsolete earlier than they will be recorded into the machine.

The most promising method of information input for the recording and storage in the electronic memory is input with the help of reading machines, which allowing automating the process of the reading of initial documents and coding them in the code accepted for the information system.

The problem of the automatic reading of texts is part of the general problem of the pattern recognition but has its specific character determined by the character of initial information. If with pattern recognition the term pattern implies a wide range of objects which can include besides signs, also figures, photographs, sounds, colors, etc., then with the automatic reading of tests the initial information is, basically a printed or typewritten document. This limits the range of examined objects by alphanumeric signs of different outlines, and therefore it is possible to discuss the technical embodiment of the reading machine.

In examining questions of the reliability of the operation of the reader, it is necessary to note the essential difference between requirements for the reliability of operation presented to figure-reading machines and requirements for the reliability of

automatic machines intended for the input information into the information-retrieval systems (IRS). If for the input of digital information in solving calculating problems the reliability of recognition is determined at 10^{-5}-10^{-6}, then in the recording of information in the IRS a reliability of the order of 10^{-3} is permissible. This means, for example, that with automatic language translation in two-three phrases of average length there can appear one error because of the information input device. Since in translation there is examined the context of the phrase, then such an error cannot lead to a distortion of the meaning. The same pertains to information input with automatic reviewing, information retrieval of bibliographic and factographic data and to the processing of certain forms of statistical data, when for increasing the reliability of reading and correction of possible errors the introduced context can be used.

The region of application of devices for the automatic reading of texts, naturally, is not limited only to systems of information retrieval and machine translation or to the expansion of the possibilities of the computer. It is possible to use these devices, for example, for the automatic composition of manuscripts, automatic sorting of letters, identification of hieroglyphs, helping the blind to read books, etc.

The flow literature increasing from year to year, which is devoted to the problem of automatic pattern identification, leaves no doubt as to the extraordinary urgency of this problem. Dozens of organizations in the Soviet Union and more than forty of the largest firms and organizations abroad are engaged in the solution of theoretical aspects of this problem and questions of the practical realization recognition devices.

In the United States the development of questions of automatic pattern recognition is being studied by about 20 of the largest industrial amalgamations; scientific research works in this field are being conducted in many universities and institutes of the

automatic machines intended for the input information into the information-retrieval systems (IRS). If for the input of digital information in solving calculating problems the reliability of recognition is determined at 10^{-5}-10^{-6}, then in the recording of information in the IRS a reliability of the order of 10^{-3} is permissible. This means, for example, that with automatic language translation in two-three phrases of average length there can appear one error because of the information input device. Since in translation there is examined the context of the phrase, then such an error cannot lead to a distortion of the meaning. The same pertains to information input with automatic reviewing, information retrieval of bibliographic and factographic data and to the processing of certain forms of statistical data, when for increasing the reliability of reading and correction of possible errors the introduced context can be used.

The region of application of devices for the automatic reading of texts, naturally, is not limited only to systems of information retrieval and machine translation or to the expansion of the possibilities of the computer. It is possible to use these devices, for example, for the automatic composition of manuscripts, automatic sorting of letters, identification of hieroglyphs, helping the blind to read books, etc.

The flow literature increasing from year to year, which is devoted to the problem of automatic pattern identification, leaves no doubt as to the extraordinary urgency of this problem. Dozens of organizations in the Soviet Union and more than forty of the largest firms and organizations abroad are engaged in the solution of theoretical aspects of this problem and questions of the practical realization recognition devices.

In the United States the development of questions of automatic pattern recognition is being studied by about 20 of the largest industrial amalgamations; scientific research works in this field are being conducted in many universities and institutes of the

country. Thus, for example, the firms of Philco and Sylvania, in collaboration with the U.S. Postal Department, are working on a reading machine for sorting postal correspondence; the firm Baird Atomic is developing devices for the translation of documents from the Russian language into English and uses the system of optical readout of information data from photographic film with the output of results of recognition on magnetic tape, the firm IBM builds optical scanners IBM-1418 and others, which reads digits from documents printed by digit-printing machines of this firm.

Works on the creation of specialized devices for recognizing signs are being conducted also in the firms Burroughs, Rabinon Engineering Co., Budd Electronics, Panoramic Research Inc., National Cash Register Co., Radio Corporation of America, General Electric, Recognition Equipment, and in others.

Of great interest are theoretical researches in the field of recognizing signs, which are being conducted both in industrial firms and in scientific establishments. Thus, for example, in the United States in the Aeronautical Laboratory at Cornell University there was given new direction to the development of works on pattern recognition; there is being proposed and developed the theory of perceptrons. At the Massachussets Institute of Technology works are being developed widely on the creation of principles instruction to recognition and modelling programs of instruction on the computer. In the universities in California, Michigan, Washington, New York, and others there is being developed a number of mathematical models, analog devices, systems, and methods of recognition.

In England interesting theoretical researches in the field of the recognition of signs are being conducted in London, Manchester, and Birmingham universities and also at the National Physics Laboratory. Specialized reading machines are being developed by firms Solartron and EMI Electronics.

In the Soviet Union theoretical and applied works on the problem of pattern recognition are being conducted successfully in scientific

research institutes of the Academy of Sciences – Institute of Problems of the Transmission of Information, Institute of Automation and Telemechanics and Kiev Institute of Cybernetics, Institute of Precision Mechanics and Computer Technology, All-Union Institute of Scientific and Technical Information and in a number of other organizations. There are being developed algorithms of recognition, methods of statistical processing of initial sets of signs, a probabilistic approach to the investigation of signs, questions of noise suppression and other questions appearing in the solution of the problem of the recognition of sign sets there appear operational models of devices realizing different algorithms of recognition. Some of these models can be considered as a prototype of future devices for the automatic reading of texts.

In accordance with regions of the application of reading machines two basic types of developed devices were determined:

1) digital reading machines designed for the processing of statistical and banking information and the input of results of recognition in computers for subsequent analysis;

2) reading machines intended basically for the input of alphabetic information for resolution of problems of information retrieval automatic reviewing, automatic translation and other problems of information service.

A characteristic peculiarity of the development of works on the creation of digital reading machines is the successful carrying out in a number of countries of the reform of printing of business documents. The fact is that the initial information for digital reading machines consists of ten digits and several special signs of the alphabet. By a special reform in England, the United States and in a number of countries of the continent signs on business documents are stylized and printed by magnetic ink. This considerably simplified the problem of recognition of the sign set and permitted creating reading machines which are quite high-speed and reliable but intended for processing of only digital documentation of a definite type.

However, if the problem of the creation of digital reading machines can be considerably simplified by means of limitation of the complex of readable signs by ten digits and their forced stylization, then the problem of the creation of alphanumeric automatic machines for the input texts in computers proves to be considerably more complicated. An analysis of the texts showed that even for the very limited problem of automatic reading of journal typographical texts of some definite branch of knowledge it will be necessary to identify approximately 10-15 different types of prints differing in dimensions, style and saturation, or 500-1000 different typographical letters. If one were to consider also interferences superimposed on the sign set as defects in the printing and different noises of the system of perception, then it becomes evident how difficult it is to create an automatic machine for the reading of at least limited typographical texts.

It is natural that best way out would be the carrying out of a general reform of printing in order to limit the quantity of utilized types of prints and to unify the selected letter taking into account requirements for printing developed in the process of the creation of reading machines. It would be expedient at first even if for printing scientific and technical literature to use straight prints of the type of cut fittings and not more than two-three different types of letters. It is desirable that for selections the basic type of print converted by proportional scaling be used. All of this most significantly would simplify the problem of the creation of the reading machine for the reading of scientific and technical literature and would permit the developers to direct their efforts not towards the creation of a unique universal device but the creation of a quite reliable and high-speed economically justified automatic machine with a wide range of application.

At present the state of works on the creation of a device for automatic readout and conversion of information can be briefly characterized in the following way:

1. The problem of pattern recognition as yet is solved, basically, in a theoretical plan and will require considerable more time, while results of theoretical researches will be able to be applied in practice for the creation of an automatic machine which reads and identifies a wide range of objects (different prints, manuscripts, figures, etc.).

2. At present it is possible in practice to place and to solve the problem of the creation of an automatic machine only for the reading of a limited set of standard printed signs, taking into account definite limitations superimposed on the form of the initial document. This will allow already in the near future the carrying out of automation of input of the main mass of scientific and technical printing literature. There is already considerable progress in this direction.

One of first operational models of a reading automatic in the Soviet Union, developed at VINITI is intended for the reading of typewritten texts in the Russian and English languages and digits. This automatic machine has been operating since 1964 and is used for the carrying out a number of scientific research works in the region of simulation of processes of recognition, accumulation of statistics of automatic reading and the solution of certain problems of information service.

Results of works on the creation of reading machines at VINITI made it possible already at a given stage to develop a preliminary project of automated technological lines, which will allow considerably reducing periods of output of signal and retrospective information and eliminating a number of laborious industrial processes.

At the Institute of Cybernetics of the Academy of Sciences of the Ukraine there has been developed an automatic machine ChARS-65, which is also intended for the reading of typewritten texts. At the end of 1966 this automatic machine was assumed complete in

development, and it is possible to trust that in 1967 there will be accumulated enough statistics for the determination of parameters of this automatic machine in speed and authenticity of reading. Similar developments are being conducted in a number of other organizations (automatic machine for the reading of typewritten and manuscript digits in the Independent design office SKB) at the Vil'nyussk plant SAM [Translator's Note: Punched-card machine plant] and an automatic machine for the reading of typewritten and typographical texts).

In any case, the problem of the creation of an automatic machine for the reading of texts with a limited set of types with a reading rate of the order of 200 signs per second and an authenticity with manual punching of the punched cards can be considered solvable both theoretically and practically. Success in the matter of manufacture and introduction of these devices will depend on the presence of sufficient design and industrial base.

3. Development of the theory of scientific information led to the revision of views on the role and place of different devices in the total system of information service. If earlier the discussion centered around the creation of the automatic device for the reading and identification of texts (written information), then now one should discuss the creation of such a system of primary data processing the basic function of which would be the automatic reading of information from initial documents of any type and the conversion of this information into a form convenient for the recording and storage in systems of information service. Actually, it is impossible to limit the information nature of a document of only the text part, since sometimes the formula or figure gives the better and more graphic representation of the contents of the document. With this if for the input of written information it is necessary to solve the problem of the automatic recognition of signs and the determination of them in the accepted alphabet, then the input of drawings, figures, and in certain cases formulas (especially for chemical compounds) can be carried out by means of

readout and automatic transfer of the examined information into the memory of the machine without identification of this information, i.e., by means of conversion of the initial representation into a certain form convenient for the recording and storage in the memory of the machine.

Available developments in the region of the creation of devices for the automatic reading of information without its identification permit drawing the conclusion that the input of figures, formulas, and other graphic materials in information retrieval systems can be automated.

Thus, in the near future it is possible to calculate the creation of complex devices necessary for the automatic readout of information and conversion of it into a form convenient for the recording and storage in systems of information service.

Problem of Information Output

The problem of the creation of information-logic systems for automated "machine" processing cannot be effectively solved without the presence of a wide complex of specialized output printing devices.

Reading machines and high-speed printers are technologically arranged on diametrically opposite poles of a single process of automatic data processing. If one were to compare the information service of a certain technological closed system connected by a wide and branched network with consumers and "suppliers" of information, then it is possible to affirm that the input and output devices are those elements which to a considerable degree determine the "viability" and effectiveness of the information system on the whole and through these elements in a wide plan is carried out the interconnection and "interaction" of information subscribers with systems of "machine" data processing.

Any information system in the end issues the consumer information data where it inevitably puts the data into a certain printing form and with automatic processing with respect to definite algorithms will convert binary, formalized and coded information into printed text or a graphic representation.

Therefore, the value of the information material and timliness of its supply to the subscriber to a considerable degree depend on the operativeness and flexibility of the system of output printers, means of copying and magnifying reproduction and remote transmission of information.

For a number of purposes it is also possible to use sound, light, and other visual-auditory form of the supply of information to the consumer with the help of radio, movie, television, and means of sign indication. However, on the contemporary stage output devices for visual and auditory perception of information have only an auxiliary importance. It is possible to affirm that the most effective form and means of mutual exchange of information between the information-logic system and the subscriber is and even for a long time will predominate the method of converting information into a certain printing publication or printed representation.

Any printing-graphic representation of information issued by an information system, should be useful on the one hand for direct perception by person and on the other - for identification with the help of "reading" machines for purposes of secondary processing or the input into new stages of information systems. In connection with this, we consider that the technological process of "machine" data processing would be more accurately depicted in the form of a closed system with effective feedback between the input and output devices.

Such an interconnection will allow providing the "reversibility" of operation of the system of information service. Thus, information material issued by the output printers when necessary can be used

for direct input into the computer and the information-logic system by means of direct "identification" on the reading machine without prolonged and complicated coding of the representation on punched tapes, punched cards, magnetic tapes and other similar carriers. On contemporary stage, in order to organize the best interconnection of input and output devices, apparently, one should ensure the printing of information in a more simple and stylized form from the point of view of the outline of types useful for reliable "perception" by reading machines.

Questions of the development of input and output devices in a wider plan can be examined as part of the general problem of direct "interaction" of man with the machine and the subscriber-consumer with the automatic information service system.

In order to increase the effectiveness of the information service in the country, it is necessary in the first place to automate the preparation and output of current information publications, signal information and new forms of publications - their kind of "lightning publications" which in their operativeness would ensure the authentic expressiveness and urgency of the delivery of information to the consumer.

The practice of advanced information services in the USSR and abroad shows that this problem can be effectively solved only on the basis of means of computer technology, punch card equipment and specialized devices of data processing. High-speed printers should provide the information-logic system or computer with the possibility of direct publishing of information publications (various kinds of author's and subject indexes, bibliographic bulletins, signal information and so forth).

In existing technological processes of machine output of information publications this problem is solved, as a rule, in two basic stages: first stage - registration of output data from the computer onto intermediate carriers (photographic film, standard paper and so forth), and the second stage - reproduction of

obtained recording for the purpose of manufacturing printed forms for subsequent circulation on typographical machines.

In order to reduce to a maximum the cycle of the output of information publications, it is necessary to ensure the technological connection of the method of recording data on the output printing device with contemporary means of "operational copying machine" and methods of manufacture of printed forms.

At the All-Union Institute of Scientific and Technical Information a technological scheme of machine output of author's and subject indexes with the help of punchcard technology was realized. The tabulogram of this complex was directly reproduced with the help of the apparatus "Era" on an offset printed form and was then circulated on apparatuses of the "Rotaprint" type.

We consider that the cardinal solution of the problem of printed circulation of data obtained from the computer and output by the machine method of information can be achieved with the transfer of results of the recording of signs on an offset printed form by the electrographic method or onto a stenciled printed form by the electrospark method.

To circulate information obtained from the computer in a small quantity of copies, highly productive electrographic apparatuses of the rotary type are effective. The development and serial production of a whole series of rotary electrographic apparatuses of the "REM" type carried out in the USSR will allow providing high-speed copying and low-circulation reproduction of information recorded directly from computers and different information-logic devices. Abroad for the reproduction of blanks information obtained from computers, rotary electrographic apparatuses Copyflo-2A (of firm Rank-Xerox Ltd, England) are widely used. The figure gives a diagram of the technological cycle of "machine" output of information articles.

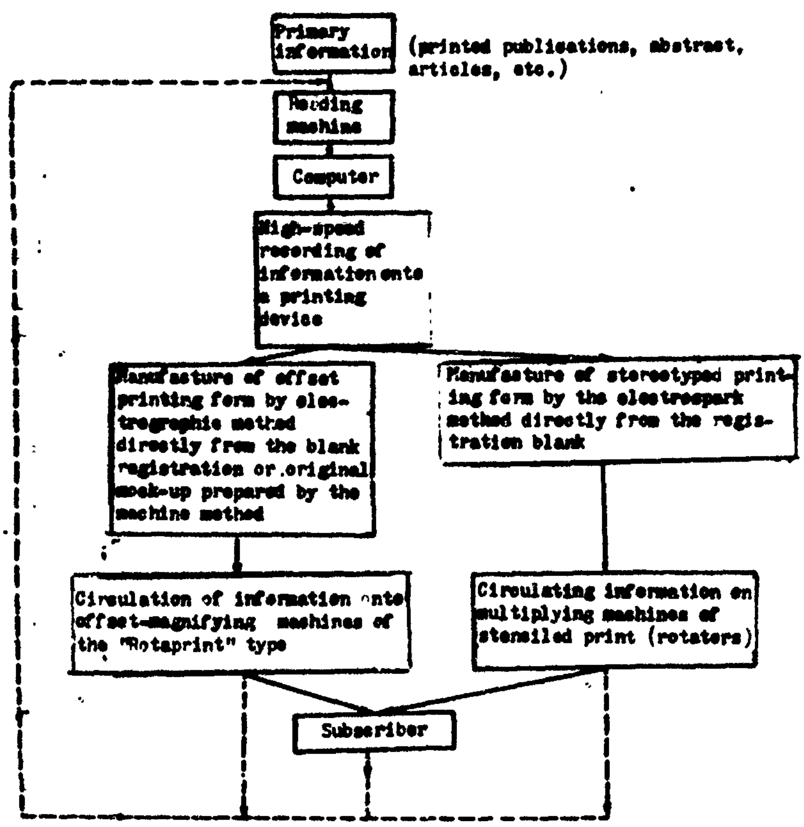
Upon recording information from the computer onto transparent carriers of the tracing paper type in a number of cases reproduction and limited circulation of data the computer by contemporary blueprint apparatuses are expedient. At present abroad there is being produced a number of specialized blueprint apparatuses intended for this purpose.

A deficiency in the examined scheme of the output of information publications is the multistage nature of the conversion of output data from the computer into circulating information publication. In order to eliminate this and increase the effectiveness of the process on the whole, we propose a fundamentally different solution to the problem of circulating output of information publications by the machine method — carry out on the output printing device registration of output data with the computer directly onto an offset printed form without the application of some additional and intermediate recording carriers. This can be carried out, for example, if one were to use as the recording part of the output printing device an electrographic apparatus of the rotary type with the transfer of the image from a selenium drum onto an offset printed form on hydrophilic paper or by electromechanical printing directly onto thin hydrophilic paper.

To ensure the effectiveness of the process of the output of information publications by means of computer technology, it is necessary to increase sharply the speed of the printing registration. An important problem and all the more "bottleneck" of computer technology becomes the increasing disproportion between the steadily increasing high-speed operation of the computer and relative low speed of the output printing devices.

The majority of Soviet and a great number of foreign commercial computers are equipped with electromechanical printing devices on a base of character wheels, lettered drums, tish-tang [Translator's Note: this term is a transliteration and could not be found], sign circuits, and so on. For machine output of information

publications abroad automatic typesetting-typewriters of the Flexowriter and Justowriter type are widely used. An important merit of machines of the Justowriter type consists in the typographical tracing of type and the possibility of ensuring print with switching off of the line. By means of a Justowriter it is possible also to carry out effectively production of "original model" of the information publication. The possibility of the automatic operation of the Flexowriter with control from punched tapes and punched cards determines its effective use in a complex with punchcard equipment. Special attachments to these machines permit carrying out remote transmission and the printing of information.



On many types of contemporary electromechanical printers very high speeds of printing for this class of machines are reached. Apparently, the world speed record for electromechanical printers

belongs to the printer of the cycloidal type developed at the Computer Center of the university in the city of Toronto (Canada), – 50 lines per second. However, in the contemporary stage it is all the more complicated to solve the problem of the sharp increase in speed of printing output data with the help of electromechanical printing devices, since in spite of the continuing growth in their high-speed operation, they, apparently, already quickly exhaust their "high-speed" possibility.

It is possible in a cardinal way to solve the problem of the creation of output printers by means of development of non-mechanical printers on the basis of a contemporary means of character indication and methods of the formation of an image.

It is possible to separate the following basic directions of the creation of nonmechanical high-speed printers into the following:

- photosetting and photorecording,
- electrophotographic (xerographic),
- electrostatic,
- electrochemical,
- electrothermal,
- electrospark,
- magnetographic,
- thermographic,
- thermoplastic and others.

An interesting process of "machine" output of information publication by means of computer technology and a specialized photosetting machine is used in the United States by the National Medical Library in a system of automated data processing in the field of medicine – Medlars.

For preparation of the complicated reference and information publication "Index Medicus" there is used a photosetting machine "Photon" of model 900 Grace (Graphic Arts Composition Equipment),

which carries out the automatic set (with control from magnetic tape of the external storage unit of a computer of the Honeywell 800 type) at a rate of 300 characters per second or 3600 words per minute. As a result the time of composition of the number of index, "Index Medicus" with a volume of 600 pages (14,000 bibliographic descriptions) including proofreading of the punched tape and circulating reproduction of the index, is 5 days.

The photosetting machine "Photon Grace" with operation for the set "Index Medicus" ensures in 16 hours of work the photosetting of more than 9,000,000 characters or 1,800,000 words!

At the All-Union Institute of Scientific and Technical Information a photosetting machine operating in a complex with a computer is developed. Results of this work are discussed in a special report.

Contemporary output printers have achieved printing rates of hundreds of thousands, a million and more characters per minute. On the xerographic device S-C-5000 of firms Stromberg Carlson and Xerox Corp. (USA), the recording rate is 1 million characters per minute. Apparently, one of the highest printing rates of data from a computer is reached on an electrothermal printing device Radiation model 690, developed by firm Radiation on order of the Lawrence Radiation Laboratory (United States). The recording speed of data on this device is 62,500 characters per second or 3,750,000 characters per minute. With the width of the line at 120 characters the high-speed operation of the device is 520 lines per second or 31,250 lines per minute. The quantity of character information recorded on the device Radiation in 1 min is equivalent to 30-50 copys of full issues of the newspaper "Pravda."

Certain American developers of high-speed printers based on the electrostatic principle consider that in the next few years in principle there can be achieved such speed of electrostatic registration which will allow recording in 1 min character information equivalent to 500 full issues of the newspaper "New York Times."

A model of a high-speed printer developed in the All-Union Institute of Scientific and Technical Information, with recording on electrothermal paper provides one of the highest speeds of data recording in the USSR - 300 lines per 1 s. Results of this work are also examined in a special report.

Thus, already on the contemporary stage it is theoretically possible to develop printers whose recording speed will be commensurable with the high-speed operation of a computer. However, it would be incorrect to consider that the simple increase in recording speed on the output device will mean full solution of the problem of data output with a computer. This question is considerably more complicated.

In the plan of "machine" output of information publications one should discuss not only the printer itself but also the operativeness, duration and economic effectiveness of the whole technological cycle of data processing. In a number of cases the less high-speed output printer can appear considerably more effective than the high-speed and ultrahigh-speed devices if the method, form and quality of recording on this device are better "inscribed" into the subsequent technology of the output of information publication. We would like once again to emphasize that by itself the increase in speed of operation of the output printer and the "pulling" of this speed to a level close to the high-speed computer operation still does not completely solve the problem of high-speed outlet.

With an increase in speeds of printers (which, as a rule, is conjugate with the sharp increase in cost and complexity of the printer[1]) it is necessary to solve a number of problems of optimum organization of the work of complex of the computer and printer. We consider that for "clearing" of limited masses of

---

[1]The cost of an experimental model of high-speed printer Radiation model 690 is $400,000.

information from the computer onto print with long periods between delivery of data, in many cases it is economically inexpedient to use high-speed and ultrahigh-speed printers.

The application of high-speed and ultrahigh-speed printers is most effective, as a rule, in the recording of great masses of information. In connection with this the direct connection of the computer with high-speed and ultrahigh-speed printers is always justified. In many cases it is more expedient to create between the computer and printer an intermediate buffer – a storage unit of information, which will be more uniform to print considerable masses of information.

The increase in speed of recording data on a printer raises the question of more complete, prolonged and continuous loading of the printer, i.e., increase in coefficient of its use. With the corresponding organization the combination of one high-speed printer with several computers is possible if the clearing of data on a printer is produced successfully with the purpose of a more complete and uniform work of the latter. Such a variant can be effectively used and realized in the creation of a single system of information service in the country and for the organization of a single network of computer centers, which carry out the overall control, planning and calculation of technical-economic data in the system of the national economy. In this case the branched network of information and computer centers should have a remote information-carrying system, and, accordingly, there should be ensured direct output from the computer through a certain communication channel onto printers installed near the subscriber. With high-speed and ultrahigh-speed printers it is possible to carry out centralized collection, processing and printing of great masses of information proceeding from different computer centers along the communication channel to one or several central computer centers equipped with high-speed and ultrahigh-speed printers able to "perceive" for printing huge masses of continuously proceeding information. In this case the utilization factor of the printers will be the highest, and the effect of high-speed and ultrahigh-speed recording the most perceptible.

ORGANIZATION OF AN INFORMATION SYSTEM BASED ON THE COMPUTER "DNEPR"

A. I. Nikitin and G. Ya. Mashbits

The basis for realization of contemporary information systems is electronic computers and their software. Requirements which are presented by the information system to the machine (and its software) lead basically to the following:

- the machine should solve the parallel exchange of information along many channels;
- the maximum service time of the interrogation in the system should be limited in accordance with requirements of the effectiveness of the system;
- the machine should allow the application in the system of different information languages, and the translators from these languages should be compact and give programs effective in high-speed operation and memory.

In a report there is described one of the possible variants of the organization of exchange in the operational system developed for the information control system (IUS) as a central link of which the computer "Dnepr" is used.

Basic parameters of the computer are the following:

- high-speed operation – 20,000 additions per second (on the average);
- system of information representation – character (8 bits per character);
- volume of IMU – up to 32,000 words (32-bit);

- volume PMU – up to 32,000 words
- volume of storage units on magnetic tapes – up to $8 \cdot 10^6$ words;
- system of interruption – on 32,000 signals;
- input device – from punched tape (photoinput);
- perforator – tape;

The machine also includes alphanumeric printers [ATsPU] (АЦПУ) teletypes (several tens) and the operator console on the basis of a typewriter (PO).

External devices are connected to the immediate-access memory unit through a special device of exchange (UO), which allows conducting in parallel work with several devices.

Algorithms of exchange were developed on the basis of the following requirements for the IUS:

- system periodically solves one (or several) problems of technological control, simulating control on a real-time scale;
- at any moment of time there is allowed access on the part of the subscribers of the information system for information material and information input from work stations according to the change in masses of information and economic data;
- in free time intervals the computer solves the calculating problems.

Organization of exchange. Requirements of high effectiveness of the computer operation place in a new fashion the question of the organization of exchange by information between man and machine. The traditional work of the operator on the panel of the machine becomes impermissible, since for all three aforementioned basic types of problems the exchange of information occurs when the machine, in general, is already occupied in the solution of other problems.

In accordance with the assignment and parameters of IUS 3 basic conditions of the exchange of information operator-machine are accepted:

- conditions of information system,
- conditions of the change in masses,
- conditions of the calculation system.

The possibility of the operation of the system in these conditions is ensured by the program-dispatcher in the presence of a corresponding complex of external devices and library of programs.

The general scheme of the exchange of information in the man-machine system is shown on Fig. 1. Man and machine are united with each other by several information channels, part of which serves for the flow of information in the direction of man-machine, part for opposite directions, and the remaining for both directions. The exchange of information occurs by separate performances, starting as a rule, on the initiative of man (operator). The exception consists of cases when the machine issues information about short-duration failures or refusals - emergency information. In this case the performance is on the initiative of the machine.

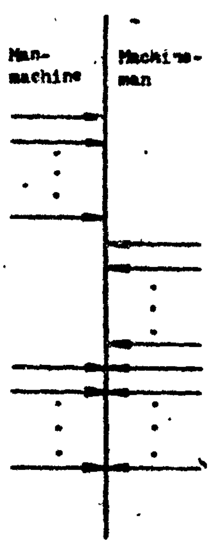


Fig. 1. General diagram of exchange in the man-machine system.

Each performance, in general, consists of several elementary performances - exchanges of messages (or simply exchanges). The number of exchanges in a performance on the General-Purpose Computer System (UVS) "Dnepr" varies from two to four and depends on conditions

in which the system operates. Each exchange consists of two messages - message of man (operator) and message of the machine. We will designate the message of the operator by letters a, b, c, and d and the corresponding return messages of the machine by the same letters with primes - a', b', c', d'.

The first exchange serves for setting the connection between the operator and the machine, and the last exchange is an exchange of basic information: for example, for the case of the solution of the calculating problem - this is the input of the program and initial data operator and output of solution by the machine. Intermediate exchanges (if they exist) serve for the preliminary exchange of working information, enabling the machine to be adjusted for a definite form of work and to issue to the operator instructions on the order of input, external devices for input and output, etc.

All messages (except for the last two) are shaped on a special (operation) language and are called reports. The operation language completely reflects the possibility of the system with respect to possible conditions of the exchange of information between the operator and machine, and its flexibility to considerable degree is governed by the effectiveness of use of the system. In general the operation language consists of two parts: a set of reports on the service of problems and a set of reports on the service of the machine. The recording of reports obeys definite syntactic rules, and in most cases it is considerably simpler than rules for universal algorithmic languages.

For UVS there is developed an operation language (Operkod), which is one of the component parts of the software system. Diagrams of the exchange for different conditions are shown in Fig. 2.

As can be seen from Fig. 2, in conditions of the calculating system performance starts with message a through the push button interrogation. The return message a' is issued by the machine on the operator console (PO) and consists in the printing of the * sign.

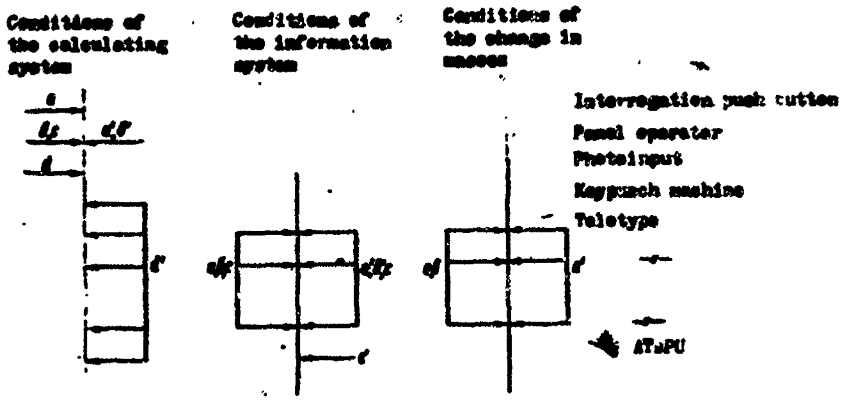


Fig. 2. Diagram of the operator-machine exchange of the UVS "Dnepr-2."

Message b is produced by the operator through the PO; in it there is indicated the priority with which the problem should be taken and solved. For calculating problems there are three level of priority: 1, 2, and 3 (in order of decrease). Message b' completely coincides with a'.

Message c (also introduced through the PO) is the certificate of the problem in which there is indicated the name of the problem, the code appropriated to it for the solution, information on input-output devices necessary for its solution and a number of arguments for the program. In content the certificate is similar for the operator of access to the procedure in ALGOL-60 language. The certificate of the problem constitutes one of the basic syntactic constructions of the operational language. Reporting by call of the problem has a form to which it is easy to lead any interrogations into the information system written on information language.

Message c' is absent, since the machine itself from the photoinput introduces a program (package d) which should be by this time prepared by operator for input. Message d' is a solution to the problem and is issued through any of the teletypes or ATsPU.

In conditions of the information system the whole performance of exchange occurs through teletypes with the exception of the answer of information system (message c'), which in the case a large volume of information is printed on the ATsPU. The form and content of the messages are the following:

Message	Form of Message	Content
a	?	Interrogation at input
a'	=	Solution on input
b	1 or 2	Priority
b'	=	Solution at next input
c		Interrogation in information language
c'		Answer of information system in a language

Conditions of the change in masses is characterized by a shorter performance of the exchange, since the preliminary priority input is absent: the operator always can introduce his data into the machine, and furthermore messages c and c' are absent.

The form and content of the messages are the following:

Messages	Form of message	Content
a	?	Interrogation at input
a'	=	Solution at input
b		Information for change in masses

DEVELOPMENT OF A DICTIONARY AND ALGORITHM FOR THE SYSTEM OF AUTOMATIC DESCRIPTOR INDEXING OF PATENT DOCUMENTATION IN THE ENGLISH LANGUAGE

A. G. Gerasimov, Yu. V. Girshchberg, and V. Z. Shenderov

For systems of patent retrieval there is great urgency in the problem of the creation of multilingual descriptor IRE, which should ensure the possibility of the retrieval of patent documentation in various languages. It is expedient to organize descriptor indexing of patent literature in the original language without preliminary translation of the indexing text into the Russian language.

A necessary condition of this is the development of a multilingual descriptor dictionary in which each idea would be adequately represented in appropriate natural languages. It is necessary simultaneously to consider that interrogations on which access to IRE will occur are formulated in the Russian language.

At present in the Central Scientific Research Institute of Patent Information and Economic Research (TsNIIPI) there has been started the development of an experimental system of automatic indexing of patent texts on the subject the "Internal-Combustion Engines." In connection with the fact that about 50% of the total volume of patent information is published in the English language, in the first stages it is expedient to automate the indexing of patent documentation in the English language.

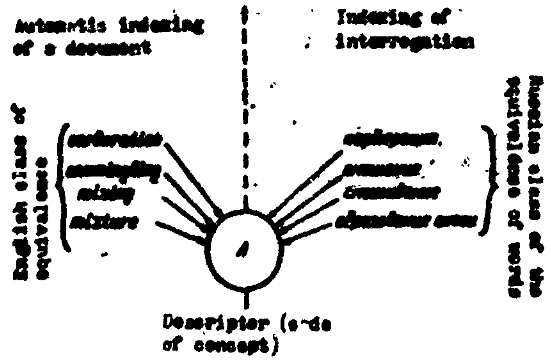
The developed system is intended for automation of the process of the descriptor indexing and, furthermore, certain operations on the organization of the descriptor retrieval system.

The main problem is an experimental check of the possibility (and expediency) of automation of a majority of stages of the processing of patent documentation for the purpose of its input into the automated retrieval system.

System is intended, basically, for the processing of signal patent documentation in the English language having an abstract character. One of the basic forms of such documentation consists of publications from the American patent weekly "Official Gazzette." However, the system will be tested in other forms of documentation: publications from the English patent journal "Abridgements of Specification" and also full patent descriptions in the English language.

As a result of work of the algorithm of indexing, in the computer memory there is recorded a set of descriptor codes correlated with those basic concepts which are contained in the text and reflected in the dictionary of indexing. In the carrying out of retrieval there is similarly indexed the retrieval interrogation and established the identity of concepts of the text of the document in the English language and concepts of interrogation expressed in the Russian language.

From this it follows that one of the basic stages of the development of an experimental bilingual IRS is the separating of classes of equivalence (synonymous rows) in the English and Russian languages and the establishing of a single-valued correspondence between them; this can be illustrated by the example given on the figure:



Functional diagram indexing.

When in the text there is encountered any of the terms[1] of the English class of equivalence shown in the left part of the figure, it will be replaced by the descriptor "A." Analogously in the retrieval interrogation any of the terms of the Russian class of equivalence, shown in the right part of the figure, will be replaced by the same descriptor "A."

With establishment of conformity between classes of equivalence of various language systems two ways are possible:

- independent development of the English and Russian part of the descriptor dictionary and subsequent establishment of the conformity between the English and Russian by synonymous rows;
- initial development of the Russian part of the dictionary of indexing and its subsequent translation into the English language.

The first way ensures full reversibility of the system of indexing, and retrieval interrogations in this case can be formulated both in the Russian and English languages. However, due to ambiguous articulation in reality by two language systems the establishment of conformity between the independently developed Russian and English classes of equivalence is hampered. Therefore, it was decided to develop first the Russian part of the dictionary of indexing and then use it in the development of the English part.

Initially a glossary of key words and work combinations covering the whole combination of concepts of the selected subject region was composed. In the composition of the glossary texts of documents were used directly, and the majority of them constitutes abstracts of patents from the Journal of Abstracts (RZh) and translations of publications of foreign patent journals on engine building.

Excluded from the glossary were very general terms (within limits of the given subject region), utilized quite frequently but not having great information value. These pertain to terms of the type "engine," "mechanism," "instrument," and so forth.

In the selection of terms for the glossary special attention was given to the use of word combinations. The introduction of terms and word combinations into the information retrieval language, based on one of the natural languages, is quite regular and justified, since the minimum semantic unit of the natural language has a binary character, including in itself identifying and differentiating components (5). Thus, for example, the lexical unit[2] "kolenchatyy val" (crankshaft) is characterized, first, by the fact that it is identified with the concept "val" (shaft) and secondly, by the fact that it is contrasted to it with the help of the criterion (modifier) "kolenchatyy" (crank). The ability of certain word combinations to convolution with preservation of the meaning – "kolenchatyy val" – "kolenval" (crankshaft) and "avtomobil'noye maslo" – "avtol" (lubricating oil), is one more confirmation of this fact. Practically any term can be presented in the form of a combination of identifying and differentiating components (5).

A small part of word combinations is the idiom, the separating in which identifying and differentiating criteria is hampered (for example, revolutions: "mertvaya tochka" (dead point) and "kholostoy khod" (idling) – the meaning of these word combinations cannot be transferred by a combination of meanings of composite components).

Separate components of word combinations and such grammatical means of information retrieval languages as indicators of the role fulfill the often adequate functions with respect to the transmission of textual relations. Indicators of the role can be examined as standard sets of differentiating components of word combinations. For example, in word combinations using the word "sistema" (system): "sistema podachi" (system of supply), "sistema zazhiganiya" (system of ignition), "sistema puska" (system of starting), etc., the differentiating components "podacha" (supply), "zazhiganiye" (ignition) and "pusk" (starting) fulfill the function of indicators of the role.

The most widespread are indicators of the role in information retrieval languages for chemistry and metallurgy where the structure of the documents is quite uniform, inasmuch as the majority of them describes methods of the obtaining of substances and technological processes.

At the Institute of Russian Language (IPYa), of the subject chosen by us it would be most convenient to use indicators of the role in the indexing of documents containing a description of working processes and functional elements of internal combustion engines. However, the overwhelming majority of the patents contains a description of separate constructive improvements of the engine. In descriptions of subassemblies and components of the constructions of machines, separate elements are found among each other in such complex relations that it is difficult to express these relations with the help of a limited set of indicators of the role. It is possible that in information languages on engine building and other areas of machine building the transfer function of textual relations between indexing concepts can be best carried out with the help of terms and word combinations.

It is necessary to note also that the introduction of word combinations in considerable measure facilitates the solution to the problem of ambiguity causing the greatest difficulties with translation of the descriptor dictionary into the English language.

Everything said earlier permits assuming that the introduction of a large quantity of word combinations (about 60% of the vocabulary of the glossary), although it considerably expands dimensions of the dictionary, is fully expedient, inasmuch as it permits most fully and accurately reflecting the whole system of concepts and relations of the indexed subject area. In general one can assume that the majority of concepts of the information retrieval language is expressed by terminologic nominal word combinations and only in separate cases of single words.

In the second stage of the work on the creation of a descriptor dictionary separation is produced of the whole initially obtained glossary into classes of equivalence. The utilized criterion of the unification of terms into classes of equivalence can be formulated in the following way: into the general class of equivalence there are united terms corresponding to concepts coinciding in the greater part of their essential criteria. From the point of view of the coincidence of essential criteria such concepts can be called synonymous within the bounds of the given information retrieval system. Thus, the terms "rabochaya smes'" (working mixture), "goryuchaya smes'" (fuel mixture), and "toplivovozdushnaya smes'" (fuel-air mixture) for the given IRS designate synonymous concepts and can be united into one class of equivalence.

The solution of the question about the unification in general classes of equivalence of terms referred to concepts found in relationships of subordinations is rather complicated. Such terms pertain to various classes of equivalence when in the subordinatec concept there is contained a combination of such criteria which in the subordinated concept must be separated. Thus, for example, the term "injection," subordinated to the term "inlet," is separated into a separate class of equivalence. If the subordinated term is not a special concept with respect to the subordinating term, then it is not separated to a special class of equivalence. Thus, the term "vstavka" (insert) is included in one class of equivalence with the term "vkladysh" (bushing).

One of the most complicated problems which must be solved in the formation of classes of equivalence is the removal of the ambiguity of terms. It is necessary to note that in the initial glossary the quantity of such terms is very small owing to its narrow subject. The removal of ambiguity is, basically, attained by realization of one of the values of the polysemantic term in word combinations where it is present as an identifying component. Thus, for example, the polysemantic term "zaryad" (charge) was not included in any of the classes of equivalence, but word combinations with this term "elektricheskiy zaryad" (electrical charge), "zaryad topliva" (charge of fuel) and so forth were used.

As a result of the unification of the whole lexical composition of the glossary into synonymous rows from 1000 key words and word combinations, 870 classes of equivalence were formed.

The Russian part of the descriptor dictionary served as the initial material in the development of the English part of the dictionary of indexing. Let us note once again that in the development of bilingual IRS obviously, it is inexpedient at the same time to develop a glossary and separate the classes of equivalence of terms independently into the Russian and English languages simultaneously. Besides the fact that for this there must be work of specialists knowing the English language so well in order to be able to formulate into it all concepts of the given subject area, considerable difficulties, apparently, should cause the subsequent setting of a one-to-one correspondence between synonymous rows of two language systems - English and Russian.

Therefore, we divide the English vocabulary into classes of equivalence on the basis of the Russian part of the dictionary of indexing already created by this time.

Here we encounter two contradictory requirements.

On the one hand, in the translation of each of the Russian terms into the English language it is necessary to consider the maximum quantity of English equivalents in order to ensure sufficient

completeness and avoid omissions of certain terms with indexing; on the other hand, it is necessary to eliminate ambiguity of English translations of a given term in order to ensure the satisfactory accuracy of indexing.

By ambiguity we mean the entry of a term in several different classes of equivalence. As one of the means of eliminating ambiguity of English terms it is expedient to use word combinations, since translations of word combinations are, as a rule, simple. For example, the word "cam" has meanings "kulachok," "kulak," "ekstsentrik," "vystup," "zub," "kulachkovaya shayba," etc.; however, the meaning of the word combination "cam follower - tolkatel'" is determined quite clearly.

We try to solve the problem of classing the polysemantic English term to the definite class of equivalence on a lexicographic level by analyzing all of its dictionary meanings.

The following sequence of actions in the translation of the descriptor dictionary into the English language was accepted:

1. Selection from the descriptor dictionary of all terms of the Russian class of equivalence.

2. Maximum complete translation each of the terms of the Russian class of equivalence into the English language with the use of Russian-English dictionaries of similar subject and translations of contexts in which the given lexical unit is encountered.

3. Reverse translation of each of the obtained English terms with the help of English-Russian dictionaries and a comparison of the obtained set of Russian meanings with initial terms of the Russian class of equivalence.

As a result of the described procedure we obtain material (in the form of a set of Russian meanings of the English term) for carrying out a comparative analysis of the Russian English part of

the descriptor dictionary. Inasmuch as the comparative analysis requires the use of a metalanguage, and also, incidentally, for the augmentation of the corresponding Russian class of equivalence.

Several variants of relationships between Russian and English lexical units are possible:

a) one-to-one correspondence; such relationships are frequent in the translation of word combinations, in this case the problem of the inclusion of the English lexical unit into the defined class of equivalence is solved without difficulties: (in the Appendix an example of this is the relationship between Russian word "razrezhennyy" and the English word "rarefied");

b) to the given English term there corresponds a certain set of values in the metalanguage, which is found with respect to the inclusion with the set of Russian lexical units of initial class of equivalence; (in the Appendix an example of this is the relationship between the Russian word "razrezheniye" and the English word "vacuum"). In this case the question of the classing of the English lexical unit to a defined class of equivalence is also simply solved;

c) to the given English term there corresponds a certain set of meanings in the metalanguage, which occurs with respect to the crossing with the set of Russian lexical units of the initial class of equivalence. The remaining part of the set of meanings of the English term can cross with several other classes of equivalence; (In the Appendix relationships between the Russian word "vakuum" and English word "void" are given as an example. In the case of ambiguity the metalanguage characteristic of the English word is compared with the whole terminologic system taking into account the basic relationships between concepts, corresponding to the examined classes of equivalence.

The initital terminologic system (Russian part of the descriptor dictionary) can be compared with data array where every class of equivalence is similar to the document, and the set of all meanings of the given English word is similar to the retrieval interrogation.

In this set there can be isolated subsets, which characterize the degree of proximity of the examined English word to some class of equivalence. In the analysis of these subsets there can be used criteria similar to criteria of semantic conformity of certain information retrieval systems [1, 2].

For this small number of cases, when the lexicographic solution of ambiguity does not lead to positive results, in the algorithm a corresponding unit is provided.

Machine Dictionary and Algorithm Indexing

The machine dictionary is divided into the following parts: dictionary of word forms, dictionary of word combinations, data array fields. The use of dictionary of word forms was found most convenient due to the weak inflectivity of the English language and uniformity of grammatical forms utilized in texts of patent descriptions. In the information cell of the dictionary of word forms the following information about the English word form is contained: the number of the class of equivalence (descriptor number) to which there pertains the key word, the criterion of homonymy and information about the entry into the word combination of the digital code of word form (if word is not a key word and is a component part of the word combination).

The dictionary of word combinations contains information about the classing of word combinations to a defined class of equivalence codes of word forms, components of word combination, and codes of indexes of word combinations.

In data array fields there is all additional information on descriptors not entering into the operational information cells, indication of ambiguity of word forms and others. Data array fields possess variable capacity which permits simply introducing necessary change into the dictionary information.

The process of automatic indexing is divided into two stages. In the first stage lexical units corresponding to key words are indexed. In the second stage terminologic word combinations are indexed.

Let us examine work of basic algorithm blocks.

1. Unit of comparison of the text of the information document with the dictionary of indexing.

Each word form of the text is compared with the word form of the dictionary of indexing and in the case of identity is replaced by an information cell.

Key word forms of the text and cells of word combination will be subject to replacement. All other word forms are omitted. For convenience of retrieval by the dictionary in word form the number of letters, l is calculated. The number of the zone of the dictionary in which it is stored, should be equal to this number l. Inside the zone the word forms are disposed in order of decreasing frequency. Frequently word forms not subject to indexing are used, they are entered into the dictionary with zero information and are erased automatically.

The basic requirement for the unit is that it should possess flexibility, allowing radically to change the content of information cells of the dictionary with minimum expenditures of technical labor.

2. In unit of the removal of ambiguity there are examined cases of lexical ambiguity requiring for their solution direct contextual analysis. The term is considered a homonym in the case when numbers of not one but several classes of equivalence are correlated with it. In the algorithm there is provided interrogation on the criterion of ambiguity to each word form of the text. If the given word is polysemous, then by the descriptor number utilized as an index of the address of the information field, the remaining descriptor numbers correlated with the word is selected. According to the scheme whose number is stored in the same information field, selection of the needed descriptor is produced.

3. Unit of key word combinations.

The work of the unit can be briefly formulated as the replacement of several English word forms united into a word combination by one information cell containing the number of class of equivalence to which the given word combination is referred.

Information about each word of the indexed text is checked in unit 3 on the index criterion of the word combination. If this criterion is absent, the following information is examined. If it contains this criterion, then according to the address, shown there, we will revert to the list of word combinations for the given index, which is contained in the dictionary of word combinations. The word combination will be identified by means of the superposition of codes of word forms of the text onto codes of word forms from the dictionary of word combinations.

The method this superposition is determined by the information of the character of the word combination: number of words, place of the index, disrupted state and maximum number of passes.

Upon identification the initial information of the text is erased, and in its place new information is placed. If identification does not occur, pass to the following word combination from socket for the given index. If not one of the word combinations of socket is found, the next information of the text is examined. To decrease the number of idle comparisons, there is introduced a check of the text on the number of elements of the revolution proceeding in succession.

Upon completion of the operation of the unit of word combinations the whole array of the text completely is indexed. By this moment it constitutes an array of information cells of two types:

- information cells containing descriptor numbers, and
- cells corresponding to word forms of the nondescriptor type.

Cells of the second type are obliterated by sending zeroes into them and packing the text. This work is conducted by unit 4.

In unit 5 formation of the retrieval form of the document is produced: all repeated descriptor numbers are removed, and the obtained array of codes registers the number of the indexed document. The retrieval form of the document is forwarded to the special array and is inverted.

Appendix. Table for the translation of Russian classes of equivalence into the English language.

	R	A	R'	A'
Descriptor	Set of key words entering into the initial class of equivalence	Set of translations of each key word into the English language	Set of reverse translations of the English term into the Russian language	Class of the equivalence of key words
	Вакуум	Vacuum	Вакуум, разрежение	Vacuum rarefaction rarefied
		Void	Полость; пустота; вакуум; пустой; безвоздушный; опоражнивать; аннулировать	
	Вакуумный	Vacuum	Вакуум, разрежение	
4	Разрежение	Vacuum	Вакуум, разрежение	
		Rarefaction	Разрежение, разжижение	
		Evacuation	Эвакуирование, выкачивание воздуха, создание вакуума, разрежение	
		Thinning	Разжижение	
	Разреженный	Rarefied	Разреженный	

Bibliography

1. Bernshteyn E. S. Formalizovannyy yazyk i kriteriy smyslovogo sootvetstviya informatsionno-poiskovoy sistemy "Pusto-Nepusto-4" (Formalized language and the criterion of semantic conformity on an information retrieval system "Empty-Not empty-4") "NTI", 1963, No. 12.

2. Bernshteyn E. S., Lakhuti D. G., Chernyavskiy V. S. O kriteriyakh otsenki ponskovykh sistem (Criteria of the appraisal of retrieval systems). "NTI" 1964, No. 3.

3. Vcherashniy R. P., Kravets L. G. Sistema obrabotki i poiska patentnoy informatsii (System of processing and retrieval of patent information). "NTI", 1966, No. 11.

4. Voprosy razrabotki mekhanizirovannoy informatsionno-poiskovoy sistemy (Questions of the development of mechanized information retrieval system) (issue 3) NIITEKhIM, 1965.

5. Martynov V. G. Kibernetika, semiotika, lingvistika (Cybernetics, semiotics, linguistics).

6. Shenderov V. Z. Eksperiment po avtomaticheskomu indeksirovaniyu publikatsiy "Ofishel gazet" (Experiment in automatic indexing of publications "Official Gazette"). "Voprosy izobretatel'stva", 1966, No. 8.

Footnotes

[1]In certain cases by term of the English class of equivalence the whole word-changing paradigm is understood.

[2]By lexical unit we mean both separate words and word combination.

QUESTIONS OF THE DYNAMIC OPTIMIZATION OF INFORMATION PROCESS IN COMPUTERS WITH HIGHLY DEVELOPED PARALLELISM

I. A. Petukhov

I. Initial Prerequisities and Formulation of the Problem

The information process (IP) is called the entire totality of interconnected actions fulfilled by the computer upon the entering into its inputs of a sequence of messages and interrogations. In the course of the IP the message will be converted and processed with subsequent recording of the proceeding information into the information field of the computer. After analysis of the proceeding interrogations[1] the information asked for is issued from the information field.

Each type of message corresponds to a certain algorithm of its analysis, conversion, processing, and retrieval of information, etc. It is assumed that algorithms are represented in a series-parallel form, that is, that along with actions are fulfilled in series, one after the other, there are such actions which can be fulfilled in parallel - for example, information retrieval in external storage unit can occur simultaneously with a rewrite of data from the intermediate storage unit into the immediate memory unit and with continuation of calculations.

We will consider that the computer in which IP takes place possesses a high degree of parallelism [1]. For the examined case the following three properties of such a machine are the most important:

1) the computer should possess a well-developed structure, i.e., have several channels of input and output data and multistage memory;
2) it is necessary that the computer have defined means of parallelism, that is, space separation and temporal separation ensured by the system of the interruption of programs;
3) devices of the computer should operate in parallel with time.

The listed properties of the computer and algorithms create favorable conditions for the execution of dynamic optimization[2] of IP realizable in the process of the computer's operation. The purpose of dynamic optimization will be considered the providing of an extreme value of certain temporary parameters introduced below.

From all the variety of questions referring to dynamic optimization (distribution of memory, optimum retrieval, combination of the fulfillment of algorithms of different types and so forth) we will examine only the orderly fulfillment of moduli[3] of algorithms. Necessity in such order appears when the intensity of flow of message is such that before the computer devices there appear turns of moduli subject to fulfillment. With this the rules by which moduli are selected from the turn have considerable influence on the temporal parameters of IP.

II. Selection and Description of the Method of Dynamic Optimization

The peculiarity of the order of moduli is the fact that it cannot be conducted beforehand but should be conducted in the course of process with each releasing of one of the computer devices. For such order it is possible to use methods of linear and dynamic programming. However, it is established [2] that in cases analogous to those examined, the cost which is obtained by solution by the shown methods frequently exceeds the saving obtained from it. A

quite simply and rapidly result is obtained with the method of suboptimization, which ensures not an optimum result but one close to the optimum. However, the available methods (see for example, [2, 3]) of suboptimization are based on the prerequisite of the known time of fulfillment of the moduli. For the examined case such a condition is unacceptable, since the time of fulfillment of moduli in the majority of computer devices is a random variable distributed according to different laws. For use in the given conditions it is possible to propose the following method.

We will consider that as a basis of the proposed method of suboptimization (we will call it simply the method of optimization) lies a key. The structure of the key is characterized by the following:

- composition of elements of the key,
- order of their location, which determines the order of calculation ion of elements of the key,
- principle of calculation of elements of the key (showing of preference to the larger or smaller value of the element of the key).

The key can be represented in such a form:

Elements of the key	ℓ_1	ℓ_2	...	ℓ_{n-1}	ℓ_n
Criterion of calculation of cells of the key	k_1	k_2	...	k_{n-1}	k_n

Elements of the key are parameters of IP located in a definite order about which will be mentioned below. The criterion of calculation of the elements of the key can have the value "0" or "1." We will consider that value "1" corresponds to the case when it is necessary to issue preference to the value of the element of the key larger in absolute magnitude and value "0," conversely, to the smaller.

So that as a result of the use of the key only one of the moduli was necessarily selected, it is necessary as the last element of the key (e_n) to use the number of the modulus corresponding to the number of the message appropriated with entering of the given message.

The principle of the use of the key will be explained by the following example.

Example. Let us assume $n = 3$, $k_1 = 0$, $k_2 = 1$, $k_3 = 0$. We will assume that in turn there are 4 moduli. The key will appear in the following way:

$$\begin{matrix} k_1 & k_2 & k_3 \\ 0 & 1 & 0 \end{matrix}$$

We will take the following values of parameters of each of the moduli (values of parameters are expressed in some relative units):

$$\begin{matrix} t_1^{(1)} = 2, & t_2^{(1)} = 3, & t_3^{(1)} = 3, \\ t_1^{(2)} = 3, & t_2^{(2)} = 4, & t_3^{(2)} = 5, \\ t_1^{(3)} = 2, & t_2^{(3)} = 3, & t_3^{(3)} = 7, \\ t_1^{(4)} = 2, & t_2^{(4)} = 2, & t_3^{(4)} = 8. \end{matrix}$$

The index in parentheses refers values of parameters to one of the four examined moduli.

Selection of the modulus subject to fulfillment will occur in the following way. In the analysis by the first elements key of the 1st, 3rd and 4th moduli will be selected. With analysis by the second element of the key, of these moduli the 1st and 3rd will remain. With analysis by the third element of the key finally for execution the 1st modulus will be selected.

If in the process of fulfillment of the 1st modulus in turn additionally not one modulus will proceed, then in the following stage, already after the analysis by the second element of the key the 3rd modulus will be selected. Under the same condition in the following section, after analysis by the first element of the key 4th modulus is selected.

III. Method and Certain Results of an Experimental Check of the Selected Method

To check the method of optimization described above $n = 4$ was accepted and the following elements were included into the

composition of the key:

- m – ordinal number of the message[4];
- l – length of message – magnitude with respect to which the time of fulfillment of moduli of the algorithm corresponding to the given message is in direct proportion to the dependence;
- r – connectivity of the modulus – quantity of moduli the fulfillment of which is retained by the fulfillment of the given modulus;
- d – aftereffect of the modulus – quantity of moduli, which remains up to the end of fulfillment of the algorithm after fulfillment of the given modulus.

The meaning of the last two of the listed parameters will be explained in the example of the graph diagram of the algorithm presented on the figure. We will consider that the given graph diagram corresponds to the algorithm of analysis of the message which is represented in series-parallel form. The peculiarity of the given graph diagram is the fact that it contains only moduli-converters and does not contain moduli-identifiers [2]. Such limitation is valid, if it is considered that to each modulus-converter there corresponds one of the criteria of fulfillment of the modulus (unconditional fulfillment, fulfillment with certain probability, fulfillment under the condition of fulfillment of the preceding modulus and so forth). Parameters r and d of moduli of the algorithm, the graph diagram of which is depicted on the figure, are given in Table 1 (with determination of values of index d the highest possible quantity of moduli was considered).

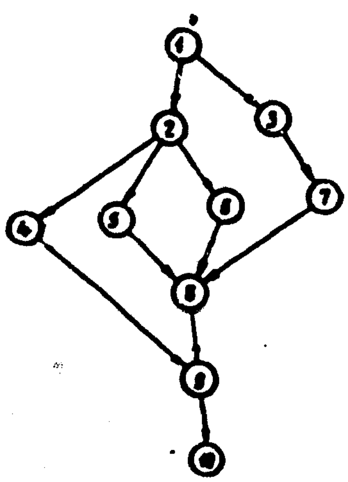


Fig. Graph diagram of the algorithm.

Table 1.

Parameters	No. of modulus									
	1	2	3	4	5	6	7	8	9	10
r	2	3	1	1	1	1	1	1	1	1
d	9	6	4	2	3	3	3	2	1	0

As was already mentioned above, in last place in the sequence of elements of the key there should be placed the modulus, i.e , $\ell_{\text{ц}} = m$. To determine the order of location of the remaining elements of the key, the following method was accepted.

As IP time parameters there was accepted the mathematical expectation (M) and dispersion (σ^2) of the time of stay of messages in the computer and also values of the function

$$t_{max} = f(\lambda),$$

where λ – intensity of input flow of the message; t_{max} – maximum time of stay of the message in the computer.

Let us explain the function.

In the course of IP in the computer every value of intensity of input flow (λ) with a defined relationship between types of messages and certain organization of IP corresponds to a certain maximum time of stay of the message in the computer (t_{max}). Thus, the process is characterized by a totality of pairs, (λ, t_{max}) or points of conformity, which form the limit of conformity (absolute or with defined probability).

We will consider L to be the measure of conformity. Here

$$L = P(t_{\text{пс}} < t_{max}),$$

where $t_{\text{пс}}$ – time of stay in the computer of an arbitrary message. That is, in the case when the measure of conformity is equal to L,

λL of the requirements are found in the computer, and time $t_{nc} \geq t_{max}$. When L = 1 there occurs absolute conformity, and when L < 1 – conformity with a certain probability.

For each element of the key, the degree of its influence on time parameters of IP was determined. To estimate the degree of influence there was acdepted the optimization factor

$$k = \frac{A_{исх} - A_{ср}}{A_{исх}},$$

where $A_{ИСХ}$ – initial magnitude of the time parameter of IP, which corresponds to the fulfillment of moduli in the order of entry, i.e., without optimization; A_{cp} – comparable magnitude of time parameter of IP, which corresponds to the case of optimization by key.

The degree of the effect of elements of the key on time parameters was determined by means of IP modeling. With this the different organization of IP and different relationships between types of messages corresponded to different values of $\kappa_{ОПТ}$. The revealed general qualitative regularities are represented in Table 2.

Naturally, only that criterion of the calculation for which $\kappa_{ОПТ} > 0$ of interest to us. From Table 2 it is clear that the ambiguity in the determination of the criterion of calculation of elements of the key appears only with respect to the number of the modulus (m), and, consequently, in this case the solution on the value of criterion of calculation must be taken in reference to concrete requirements for time parameters.

Elements of the key in sequence of elements cells of the key were disposed in the order of decrease of $\kappa_{ОПТ}$.

In accordance with the expounded method in the modeling of IP the following key was obtained:

$$\begin{matrix} l & d & r & m \\ 0 & 0 & 1 & 0 \end{matrix}$$

Table 2.

Elements of the key with criteria of their calcu-lation		The parameters of IP				
		m	[illegible]	[illegible]		
				[illegible]	[illegible]	[illegible]
		[illegible]				
m	$k=1$	>0	<0	<0	>0	>0
	$k=0$	=0	=0	=0	=0	=0
l	$k=1$	<0	<0	<0	<0	<0
	$k=0$	>0	>0	>0	>0	>0
r	$k=1$	>0	>0	>0	>0	>0
	$k=0$	<0	<0	<0	<0	<0
d	$k=1$	<0	<0	<0	<0	<0
	$k=0$	>0	>0	>0	>0	>0

The use of such key ensures the value of the optimization factor within limits of $\kappa_{\text{ОПТ}} = 0.2\text{-}0.8$. An increase in value $\kappa_{\text{ОПТ}}$ occurs with an increase in intensity of flow of the messages.

Cases are known when messages are given different categories of preference (priorities). The category of preference should be set in the first place in the key with the corresponding criterion of calculation. The introduction of the category of preference in general somewhat lowers the effect of the application of the key.

It should be indicated that the given key is not universal and can be changed together with a change in parameters of algorithms.

Conclusion

With branched algorithms fulfilled by the computer with highly developed parallelism, favorable conditions for optimization of IP, carried out in the process of the computer operation, are created. One of directions of such optimization is the order of fulfillment of sections of algorithms (moduli). The necessity in order appears when at a certain intensity of the input flow of messages in front of the computer devices, lines of moduli appear.

The method described permits, with the help of a simple procedure, selecting from the line moduli subject to fulfillment. For selection there is used a key which constitutes the totality of parameters of moduli located in defined order and equipped with criteria of calculation. As a result of analysis by the key from the line only one modulus is always selected.

Use of the key permits reducing the time of stay of messages in the computer by two and more times.

It is very important that in the described method, in contrast to other known methods of the order moduli, it is not necessary to know preliminarily the time of fulfillment of the moduli. In order to use the method in real conditions, it is sometimes necessary to consider the peculiarities of structure of the specific computer.

Bibliography

1. Rayl B. P. Mul'tiprogrammnaya obrabotka dannykh (Multiprogramming of data processing). Kiberneticheskiy sbornik, vyp. No. 5, Izd-vo "Mir", M., 1962.

2. Shvarts Yu. S. Avtomaticheskiy protsess uporyadocheniya moduley i yego prilozheniye k parallel'nomu programmirovaniyu (Automatic process of the order of moduli and its application to parallel programming). Kiberneticheskiy sbornik, vyp 9, Izd-vo "Mir", M., 1964.

3. Dinamicheskaya model' optimal'noy zagruzki melkoseriynykh predpriyatiy. Ekspressinformatsiya (Dynamic model of optimum loading of small-scale undertakings. Rapid analysis). Tekhnicheskaya kibernetika, No. 19, 1965.

Footnotes

[1]In the subsequent account the concept "message" unites both messages and interrogations.

[2]The term "dynamic optimization" of IP is selected in contrast to "static optimization" which is fulfilled at the stage of preparation of algorithms.

[3]The modulus is equivalent the part of the algorithm which can be expressed by a final number of instructions of the computer and carried out by one of its devices.

[4]It is assumed that such an ordinal number is given to each message proceeding to the computer. This number corresponds to each modulus of the algorithm of the processing of the given message.

DATA HANDLING PAGE

01-ACCESSION NO. / 98-DOCUMENT LOC: TT9501058	39-TOPIC TAGS: information storage and retrieval, data processing system, data processing personnel, computer input unit, reading machine, pattern recognition, digital computer, computer output unit
09-TITLE: QUESTIONS OF THE AUTOMATION OF INPUT, OUTPUT, AND DATA PROCESSING ON COMPUTERS IN SYSTEMS OF INFORMATION SERVICE	
47-SUBJECT AREA: 09, 05	

42-AUTHOR/CO-AUTHORS: MAMONTOV, O. V. ; 16-AVRUKH, M. L.; 16-KAL'MANSON, V. A.	10-DATE OF INFO: -----67
43-SOURCE: VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)	68-DOCUMENT NO.: FTD-MT-24-133-69
	69-PROJECT NO.: 6050205

63-SECURITY AND DOWNGRADING INFORMATION	64-CONTROL MARKINGS	97-HEADER CLASN
UNCL, 0	NONE	UNCL

76-REEL FRAME NO.	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA	NO OF PAGES
1889 1281			UR, US	23
CONTRACT NO.	**X REF ACC. NO.**	**PUBLISHING DATE**	**TYPE PRODUCT**	**REVISION FREQ**
	65-	94-00	TRANSLATION	NONE

STEP NO.	ACCESSION NO.
02-UR/0000/67/004/000/0025/0038	

ABSTRACT

(U) This article deals with questions of the automation of input, output and data processing on computers in systems of information service. The authors discuss peculiarities of applied computers and processed information, general characteristics of processing algorithms, problem of information input and problem of information output.

DATA HANDLING PAGE

01-ACCESSION NO. / 98-DOCUMENT LOC	TT9001059
39-TOPIC TAGS	information storage and retrieval, computer language/(U)DNEPR control computer
09-TITLE	ORGANIZATION OF AN INFORMATION SYSTEM BASED ON THE COMPUTER "DNEPR"
47-SUBJECT AREA	09
42-AUTHOR/CO-AUTHORS	NIKITIN, A. I.; 16-MASHBITS, G. YA.
10-DATE OF INFO	-----67
43-SOURCE	VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)
68-DOCUMENT NO.	FTD-MT-24-133-69
69-PROJECT NO	6050205
63-SECURITY AND DOWNGRADING INFORMATION	UNCL, O
64-CONTROL MARKINGS	NONE
97-HEADER CLASN	UNCL
76-REEL FRAME NO.	1889 1282
77-SUPERSEDES	
78-CHANGES	
40-GEOGRAPHICAL AREA	UR
NO OF PAGES	6
CONTRACT NO.	
X REF ACC. NO.	65-
PUBLISHING DATE	94-00
TYPE PRODUCT	TRANSLATION
REVISION FREQ	NONE
STEP NO.	02-UR/0000/67/004/000/0203/0206
ACCESSION NO.	

ABSTRACT

(U) The organization of an information system based on the Soviet computer "DNEPR" is discussed. The basic parameters of the computer are given. The exchange of information in the man-machine system is touched upon, and a diagram is given showing this. Orig. art has: 2 figures.

DATA HANDLING PAGE

01-ACCESSION NO. / 98-DOCUMENT LOC	TT9501060
39-TOPIC TAGS	automatic document analysis, machine abstracting, information storage and retrieval
09-TITLE	DEVELOPMENT OF A DICTIONARY AND ALGORITHM FOR THE SYSTEM OF AUTOMATIC DESCRIPTOR INDEXING OF PATENT DOCUMENTATION IN THE ENGLISH LANGUAGE
47-SUBJECT AREA	09
42-AUTHOR/CO-AUTHORS	GERASIMOV, A. G.; 16-GIRSHCHBERG, YU. V.; 16-SHE[illegible]EROV, V. Z.
10-DATE OF INFO	-----67
43-SOURCE	VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)
68-DOCUMENT NO.	FTD-MT-24-133-69
69-PROJECT NO.	6050205
63-SECURITY AND DOWNGRADING INFORMATION	UNCL, 0
64-CONTROL MARKINGS	NONE
97-HEADER CLASN	UNCL
76-REEL FRAME NO.	1889 1283
77-SUPERSEDES	
78-CHANGES	
40-GEOGRAPHICAL AREA	UR
NO OF PAGES	14
CONTRACT NO.	
X REF ACC. NO.	65-
PUBLISHING DATE	94-00
TYPE PRODUCT	TRANSLATION
REVISION FREQ	NONE
STEP NO.	02-UR/0000/67/004/000/0222/0229
ACCESSION NO.	

ABSTRACT

(U) The development of a dictionary and algorithm for the system of automatic descriptor indexing of patent documentation in the English language is discussed. It is noted that the indexing of patent literature should be done in the original language without preliminary translation of the indexing text into the Russian language. A brief discussion is given of a machine dictionary and algorithm indexing.

DATA HANDLING PAGE

01-ACCESSION NO. / 98-DOCUMENT LOC	TT9501061
39-TOPIC TAGS	information storage and retrieval, computer language, data processing system
09-TITLE	QUESTIONS OF THE DYNAMIC OPTIMIZATION OF INFORMATION PROCESS IN COMPUTERS WITH HIGHLY DEVELOPED PARALLELISM
47-SUBJECT AREA	09
42-AUTHOR/CO-AUTHORS	PETUKHOV, I. A.
10-DATE OF INFO	-----67
43-SOURCE	VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)
68-DOCUMENT NO.	FTD-MT-24-133-69
69-PROJECT NO.	6050205
63-SECURITY AND DOWNGRADING INFORMATION	UNCL, O
64-CONTROL MARKINGS	NONE
97-HEADER CLASN	UNCL
76-REEL FRAME NO.	1889 1284
77-SUPERSEDES	
78-CHANGES	
40-GEOGRAPHICAL AREA	UR
NO OF PAGES	10
CONTRACT NO.	
X REF ACC. NO.	65-
PUBLISHING DATE	94-00
TYPE PRODUCT	TRANSLATION
REVISION FREQ	NONE
STEP NO.	02-UR/0000/67/004/000/0300 0305
ACCESSION NO.	

ABSTRACT

(U) This article deals with the dynamic optimization of information process in computers with highly developed parallelism. The article is divided into these sections: 1) initial prerequisites and formulation of the problem; 2) selection and description of the method of dynamic optimization; 3) method and certain results of an experimental check of the selected method. Orig. art. has: 2 tables and 3 formulas.

DATA HANDLING PAGE

01-ACCESSION NO.	98-DOCUMENT LOC	39-TOPIC TAGS
TT9501061		information storage and retrieval, computer language, data processing system

09-TITLE QUESTIONS OF THE DYNAMIC OPTIMIZATION OF INFORMATION PROCESS IN COMPUTERS WITH HIGHLY DEVELOPED PARALLELISM

47-SUBJECT AREA 09

42-AUTHOR/CO-AUTHORS	10-DATE OF INFO
PETUKHOV, I. A.	-----67

43-SOURCE	68-DOCUMENT NO.
VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)	FTD-MT-24-133-69
	69-PROJECT NO. 6050205

63-SECURITY AND DOWNGRADING INFORMATION	64-CONTROL MARKINGS	97-HEADER CLASN
UNCL, 0	NONE	UNCL

76-REEL FRAME NO.	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA	NO OF PAGES
1889 1284			UR	10
CONTRACT NO.	**X REF ACC. NO.**	**PUBLISHING DATE**	**TYPE PRODUCT**	**REVISION FREQ**
	65-	94-00	TRANSLATION	NONE

STEP NO.	ACCESSION NO.
02-UR/0000/67/004/000/0300 0305	

ABSTRACT

(U) This article deals with the dynamic optimization of information process in computers with highly developed parallelism. The article is divided into these sections: 1) initial prerequisites and formulation of the problem; 2) selection and description of the method of dynamic optimization; 3) method and certain results of an experimental check of the selected method. Orig. art. has: 2 tables and 3 formulas.